

# RECEIVED

Page 1 of 8

JUL 03 2003



1600

## RAW SEQUENCE LISTING

DATE: 07/01/2003

PATENT APPLICATION: US/09/820,843B

TIME: 09:06:11

Input Set : A:\09-820,843 Sequence Listing.txt

Output Set: N:\CRF4\07012003\I820843B.raw

3 <110> APPLICANT: Council of Scientific and Industrial Research  
 5 <120> TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE  
 6 PROTEINS USEFUL AS ANTI-INFECTIVES  
 8 <130> FILE REFERENCE: Q63915  
 C--> 10 <140> CURRENT APPLICATION NUMBER: US/09/820,843B  
 C--> 10 <141> CURRENT FILING DATE: 2001-03-30  
 10 <160> NUMBER OF SEQ ID NOS: 118  
 12 <170> SOFTWARE: PatentIn version 3.2  
 14 <210> SEQ ID NO: 1  
 15 <211> LENGTH: 51  
 16 <212> TYPE: PRT  
 17 <213> ORGANISM: C. jejuni  
 20 <220> FEATURE:  
 21 <221> NAME/KEY: misc\_feature  
 22 <223> OTHER INFORMATION: highly acidic protein  
 24 <220> FEATURE:  
 25 <221> NAME/KEY: misc\_feature  
 26 <223> OTHER INFORMATION: gi|6967728  
 28 <400> SEQUENCE: 1  
 30 Met Ala Tyr Glu Asp Glu Glu Asp Leu Asn Tyr Asp Asp Tyr Glu Asn  
 31 1 5 10 15  
 34 Glu Asp Glu Glu Tyr Pro Gln Asn His His Lys Asn Tyr Asn Tyr Asp  
 35 20 25 30  
 38 Asp Asp Asp Tyr Glu Tyr Asp Asp Asp Asn Asn Asp Asp Asp Phe Tyr  
 39 35 40 45  
 42 Glu Met Asp  
 43 50  
 46 <210> SEQ ID NO: 2  
 47 <211> LENGTH: 32  
 48 <212> TYPE: PRT  
 49 <213> ORGANISM: C. jejuni  
 52 <220> FEATURE:  
 53 <221> NAME/KEY: misc\_feature  
 54 <223> OTHER INFORMATION: small hydrophobic protein  
 56 <220> FEATURE:  
 57 <221> NAME/KEY: misc\_feature  
 58 <223> OTHER INFORMATION: gi|6969129  
 60 <400> SEQUENCE: 2  
 62 Met Thr Met Leu Asp Ile Phe Glu Ile Ile Phe Ile Thr Thr Val Val  
 63 1 5 10 15  
 66 Ile Ile Gly Phe Gly Gly Ile Val Phe Val Val Thr Lys Glu Lys Lys  
 67 20 25 30  
 70 <210> SEQ ID NO: 3

P.6

# ENTERED

## RAW SEQUENCE LISTING

DATE: 07/01/2003

PATENT APPLICATION: US/09/820,843B

TIME: 09:06:11

Input Set : A:\09-820,843 Sequence Listing.txt

Output Set: N:\CRF4\07012003\I820843B.raw

```

71 <211> LENGTH: 57
72 <212> TYPE: PRT
73 <213> ORGANISM: C. jejuni
76 <220> FEATURE:
77 <221> NAME/KEY: misc_feature
78 <223> OTHER INFORMATION: putative coiled coil protein
80 <220> FEATURE:
81 <221> NAME/KEY: misc_feature
82 <223> OTHER INFORMATION: gi|6968493
84 <400> SEQUENCE: 3
86 Met Ser Phe Glu Glu Asn Leu Lys His Ala Asn Glu Ser Leu Glu Lys
87 1 5 10 15
90 Leu Asn Asn Gln Glu Leu Ala Leu Asp Glu Ser Val Lys Ile Tyr Lys
91 20 25 30
94 Glu Gly Leu Glu Ser Ile Lys Lys Ala Arg Leu Glu Leu Glu Lys Ala
95 35 40 45
98 Lys Leu Glu Val Glu Gln Ile Asp Glu
99 50 55
102 <210> SEQ ID NO: 4
103 <211> LENGTH: 542
104 <212> TYPE: PRT
105 <213> ORGANISM: C. jejuni
108 <220> FEATURE:
109 <221> NAME/KEY: misc_feature
110 <223> OTHER INFORMATION: highly acidic protein
112 <220> FEATURE:
113 <221> NAME/KEY: misc_feature
114 <223> OTHER INFORMATION: gi|6968611
116 <400> SEQUENCE: 4
118 Met Lys Ile Leu Leu Leu Asn Glu Asn Pro Val Val Ser Arg Leu Val
119 1 5 10 15
122 Ser Leu Ser Ala Lys Lys Met Ser Tyr Asp Phe Glu Glu Leu Asn Ala
123 20 25 30
126 Tyr Ser Glu Asn Leu Gly Asn Tyr Asp Val Ile Val Val Asp Ser Asp
127 35 40 45
130 Thr Pro Ala Pro Leu Lys Ile Leu Lys Glu Lys Cys Asp Arg Leu Ile
131 50 55 60
134 Phe Leu Ala Pro Arg Asn Gln Asn Val Glu Asp Ile Asp Ala Gln Ile
135 65 70 75 80
138 Leu Gln Lys Pro Phe Leu Pro Thr Asp Phe Leu Asn Leu Leu Asn Asn
139 85 90 95
142 Lys Asp Ala Asn Lys His Thr Ser Ile Asp Leu Pro Met Leu Ser Asn
143 100 105 110
146 Asp Glu Asn Pro Tyr Ala Asp Ile Ser Leu Asp Leu Asp Asn Leu Asn
147 115 120 125
150 Leu Asp Asp Leu Pro Asp Glu Asn Ser Leu Asp Ile Asn Ser Glu Gly
151 130 135 140
154 Met Glu Asp Leu Ser Phe Asp Asp Ala Gln Asp Asp Asn Ala Asn
155 145 150 155 160

```

## RAW SEQUENCE LISTING

DATE: 07/01/2003

PATENT APPLICATION: US/09/820,843B

TIME: 09:06:11

Input Set : A:\09-820,843 Sequence Listing.txt

Output Set: N:\CRF4\07012003\I820843B.raw

```

158 Lys Thr Leu Glu Thr Gln Asn Leu Glu His Glu Thr Ile Lys Glu Gln
159                               165                               170                               175
162 Thr Gln Glu Asp Thr Gln Ile Asp Leu Asp Leu Thr Leu Glu Asp Gly
163                               180                               185                               190
166 Glu Ser Glu Lys Glu Asp Leu Ser Gln Glu His Thr Ala Leu Asp Thr
167                               195                               200                               205
170 Glu Pro Ser Leu Asp Glu Leu Asp Asp Lys Asn Asp Glu Asp Leu Glu
171                               210                               215                               220
174 Ile Lys Glu Asp Asp Lys Asn Glu Glu Ile Glu Lys Gln Glu Leu Leu
175 225                               230                               235                               240
178 Asp Asp Ser Lys Thr Asn Thr Leu Glu Met Gln Glu Glu Leu Ser Glu
179                               245                               250                               255
182 Ser Gln Asp Asp Asn Ser Asn Lys Thr Leu Glu Thr Gln Asn Leu Glu
183                               260                               265                               270
186 His Asp Asn Leu Glu Gln Glu Thr Ile Lys Glu Gln Thr Gln Glu Asp
187                               275                               280                               285
190 Thr Gln Ile Asp Leu Asp Leu Thr Leu Glu Asp Gly Glu Ser Glu Lys
191                               290                               295                               300
194 Glu Asp Leu Ser Gln Glu His Thr Ala Leu Asp Thr Glu Pro Ser Leu
195 305                               310                               315                               320
198 Asp Glu Leu Asp Asp Lys Asn Asp Glu Asp Leu Glu Asp Asn Lys Glu
199                               325                               330                               335
202 Leu Gln Ala Asn Ile Ser Asp Phe Asp Asp Leu Pro Glu Val Glu Glu
203                               340                               345                               350
206 Gln Glu Lys Glu Met Asp Phe Asp Asp Leu Pro Glu Asp Ala Glu Phe
207                               355                               360                               365
210 Leu Gly Gln Ala Lys Tyr Asn Glu Glu Ser Glu Glu Asn Leu Glu Glu
211                               370                               375                               380
214 Phe Ala Pro Val Val Glu Glu Asp Ile Gln Asp Glu Ile Asp Asp Phe
215 385                               390                               395                               400
218 Ala Ser Asn Leu Ser Thr Gln Asp Gln Ile Lys Glu Glu Leu Ala Gln
219                               405                               410                               415
222 Leu Asp Glu Leu Asp Tyr Gly Ile Asp Ser Asp Asn Ser Ser Lys Val
223                               420                               425                               430
226 Leu Glu Asp Phe Lys Asp Glu Pro Ile Leu Asp Asp Lys Glu Leu Gly
227                               435                               440                               445
230 Thr Asn Glu Glu Glu Val Val Val Pro Asn Leu Asn Ile Ser Asp Phe
231                               450                               455                               460
234 Asp Thr Leu Lys Glu Ser Asp Ile Gln Glu Ala Leu Gly Glu Glu Ile
235 465                               470                               475                               480
238 Leu Glu Lys Asn Glu Glu Pro Ile Val Ser Asp Val Thr Lys Asp Asp
239                               485                               490                               495
242 Asn Ser Glu Glu Ile Val Asn Glu Leu Ser Gln Ser Ile Ala Gly Ala
243                               500                               505                               510
246 Ile Thr Ser Ser Ile Lys Asp Asp Thr Leu Lys Ala Ala Leu Lys Gly
247                               515                               520                               525
250 Met Asn Met Asn Ile Asn Ile Asn Ile Ser Phe Lys Glu Asp
251                               530                               535                               540
254 <210> SEQ ID NO: 5

```

## RAW SEQUENCE LISTING

DATE: 07/01/2003

PATENT APPLICATION: US/09/820,843B

TIME: 09:06:11

Input Set : A:\09-820,843 Sequence Listing.txt

Output Set: N:\CRF4\07012003\I820843B.raw

```

255 <211> LENGTH: 172
256 <212> TYPE: PRT
257 <213> ORGANISM: C. pneumoniaeCWL029
260 <220> FEATURE:
261 <221> NAME/KEY: misc_feature
262 <223> OTHER INFORMATION: histone like protein 2
264 <220> FEATURE:
265 <221> NAME/KEY: misc_feature
266 <223> OTHER INFORMATION: gi|4376663
268 <400> SEQUENCE: 5
270 Met Ile Gly Ala Gln Lys Lys Gln Ser Gly Lys Lys Thr Ala Ser Arg
271 1 5 10 15
274 Ala Val Arg Lys Pro Ala Lys Lys Val Ala Ala Lys Arg Thr Val Lys
275 20 25 30
278 Lys Ala Thr Val Arg Lys Thr Ala Val Lys Lys Pro Ala Val Arg Lys
279 35 40 45
282 Thr Ala Ala Lys Lys Thr Val Ala Lys Lys Thr Thr Ala Lys Arg Thr
283 50 55 60
286 Val Arg Lys Thr Val Ala Lys Lys Pro Ala Val Lys Lys Val Ala Ala
287 65 70 75 80
290 Lys Arg Val Val Lys Lys Thr Val Ala Lys Lys Thr Thr Ala Lys Arg
291 85 90 95
294 Ala Val Arg Lys Thr Val Ala Lys Lys Pro Val Ala Arg Lys Thr Thr
295 100 105 110
298 Val Ala Lys Gly Ser Pro Lys Lys Ala Ala Ala Cys Ala Leu Ala Cys
299 115 120 125
302 His Lys Asn His Lys His Thr Ser Ser Cys Lys Arg Val Cys Ser Ser
303 130 135 140
306 Thr Ala Thr Arg Lys His Gly Ser Lys Ser Arg Val Arg Thr Ala His
307 145 150 155 160
310 Gly Trp Arg His Gln Leu Ile Lys Met Met Ser Arg
311 165 170
314 <210> SEQ ID NO: 6
315 <211> LENGTH: 63
316 <212> TYPE: PRT
317 <213> ORGANISM: C. trachomatis
320 <220> FEATURE:
321 <221> NAME/KEY: misc_feature
322 <223> OTHER INFORMATION: hypothetical protein-possible frameshift with CT593
324 <220> FEATURE:
325 <221> NAME/KEY: misc_feature
326 <223> OTHER INFORMATION: gi|3522902
328 <400> SEQUENCE: 6
330 Met Phe Thr Leu Phe Leu Cys Glu His Leu Leu Thr Asn Ile Leu Ala
331 1 5 10 15
334 Ser Ser Phe Leu Ala Lys Ser Gln Gly Phe Ile Thr Leu Val Asn Leu
335 20 25 30
338 Phe His Lys Ile Pro Gly Leu Lys Val Ile Glu Ile Thr Cys Leu Ala
339 35 40 45

```

## RAW SEQUENCE LISTING

DATE: 07/01/2003

PATENT APPLICATION: US/09/820,843B

TIME: 09:06:11

Input Set : A:\09-820,843 Sequence Listing.txt

Output Set: N:\CRF4\07012003\I820843B.raw

```

342 Leu Pro Leu Gly Ile His Ser Ile Ile Gly Phe Ser Tyr Leu Leu
343      50                      55                      60
346 <210> SEQ ID NO: 7
347 <211> LENGTH: 203
348 <212> TYPE: PRT
349 <213> ORGANISM: C. trachomatis
352 <220> FEATURE:
353 <221> NAME/KEY: misc_feature
354 <223> OTHER INFORMATION: histone like protein 2
356 <220> FEATURE:
357 <221> NAME/KEY: misc_feature
358 <223> OTHER INFORMATION: gi|3328438
360 <400> SEQUENCE: 7
362 Met Asn Met Leu Gly Val Gln Lys Lys Cys Ser Thr Arg Lys Thr Ala
363 1      5                      10                      15
366 Ala Arg Lys Thr Val Val Arg Lys Pro Ala Ala Lys Lys Thr Ala Ala
367      20                      25                      30
370 Lys Lys Ala Pro Val Arg Lys Val Ala Ala Lys Lys Thr Val Ala Arg
371      35                      40                      45
374 Lys Thr Val Ala Lys Lys Thr Val Ala Ala Arg Lys Pro Val Ala Lys
375      50                      55                      60
378 Lys Ala Thr Ala Lys Lys Ala Pro Val Arg Lys Val Ala Ala Lys Lys
379 65      70                      75                      80
382 Thr Val Ala Arg Lys Thr Val Ala Lys Lys Thr Val Ala Ala Arg Lys
383      85                      90                      95
386 Pro Val Ala Lys Lys Ala Thr Ala Lys Lys Ala Pro Val Arg Lys Ala
387      100                     105                     110
390 Val Ala Lys Lys Thr Val Ala Arg Lys Thr Val Ala Lys Lys Thr Val
391      115                     120                     125
394 Ala Ala Arg Lys Pro Val Ala Lys Arg Val Ala Ser Thr Lys Lys Ser
395      130                     135                     140
398 Ser Ile Ala Val Lys Ala Gly Val Cys Met Lys Lys His Lys His Thr
399 145      150                     155                     160
402 Ala Ala Cys Gly Arg Val Ala Ala Ser Gly Val Lys Val Cys Ala Ser
403      165                     170                     175
406 Ala Ala Lys Arg Lys Thr Asn Pro Asn Arg Ser Arg Thr Ala His Ser
407      180                     185                     190
410 Trp Arg Gln Gln Leu Met Lys Leu Val Ala Arg
411      195                     200
414 <210> SEQ ID NO: 8
415 <211> LENGTH: 372
416 <212> TYPE: PRT
417 <213> ORGANISM: H. influenzae
420 <220> FEATURE:
421 <221> NAME/KEY: misc_feature
422 <223> OTHER INFORMATION: outer membrane integrity protein (tolA)
424 <220> FEATURE:
425 <221> NAME/KEY: misc_feature
426 <223> OTHER INFORMATION: gi|1573353

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 07/01/2003  
PATENT APPLICATION:    US/09/820,843B      TIME: 09:06:12

Input Set : A:\09-820,843 Sequence Listing.txt  
Output Set: N:\CRF4\07012003\I820843B.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:32; Xaa Pos. 198

Seq#:66; Xaa Pos. 2